

Taxonomies in Support of Search

by
Heather Hedden

Taxonomy Boot Camp
Washington, DC

November 17, 2025

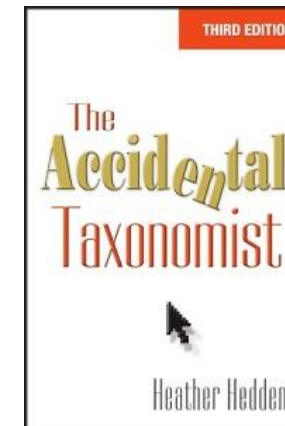


About Heather Hedden

- Independent taxonomy consultant, Hedden Information Management
- Instructor of online and corporate taxonomy courses and workshops
- Previously a taxonomy consultant in consulting firms, Enterprise Knowledge and PPC, and a contract consultant for others.
- Former taxonomy-related roles at Northern Light, Semantic Web Company, Gale/Cengage, Viziant, and First Wind.
- Author of *The Accidental Taxonomist*, 3rd ed. (2022, Information Today, Inc.)



**Hedden Information
Management**
Making information findable



Outline

- Background and Historical Trends
- Scenarios for Information Seeking
- Search on Taxonomy Terms
- Faceted Taxonomies
- Post Search Refinement
- Creating Synonyms and Search Thesauri
- Taxonomies in Support for Querying

Background

What you can do with a taxonomy



- **Search:** find content about.... (search string matches taxonomy concepts)



- **Topic browse:** explore subjects arranged in a hierarchy and then content on the subject



- **Faceted (filtering/refining) search:** find content meeting a combination of basic criteria



- **Discovery:** find other content tagged with same concepts as tagged to found content; explore broader, narrower, and (sometimes) related taxonomy topics



- **Tagging:** index content consistently so that retrieval is comprehensive and accurate



- **Content curation:** create feeds or alerts based on user profile (personalization) and taxonomy concepts tagged



- **Metadata management:** for retrieval, identification, comparison, analysis, etc.






- Combine with an ontology for even more capabilities


Historical Trends of Taxonomies and Search in User Interfaces

1. Originally, full taxonomy hierarchical browse or thesaurus alphabetical browse. Search introduced as separate from browse taxonomies. Browse *or* Search.
2. Full, large taxonomies came to be displayed less. Search added *on* taxonomy terms.
3. Search on more than just the taxonomy terms. Search on a *combination* of taxonomy terms and words in titles, headings, captions, and full text, with greater weighting on tagged taxonomy terms.
4. Faceted taxonomies – combining search and limited browsing. Especially with smaller controlled vocabularies.
5. Large taxonomies/thesauri also integrated into faceted displays. Taxonomy terms as post-search filters.
6. GenAI prompts and queries – searching with full questions, not just keywords



Historical Trends: Browse or Search

[New User? Register](#) | [Sign In](#) | [Help](#)[Get new Yahoo! Mail apps](#) |  [Mail](#) | [My Y!](#) |  [Yahoo!](#)

 **DIRECTORY**

[Search Web](#)

Yahoo! Directory [Advanced Search](#) [Suggest a Site](#)

Arts & Humanities Photography, History, Literature...	News & Media Newspapers, Radio, Weather, Blogs...
Business & Economy B2B, Finance, Shopping, Jobs...	Recreation & Sports Sports, Travel, Autos, Outdoors...
Computer & Internet Hardware, Software, Web, Games...	Reference Phone Numbers, Dictionaries, Quotes...
Education Colleges, K-12, Distance Learning...	Regional Countries, Regions, U.S. States...
Entertainment Movies, TV Shows, Music, Humor...	Science Animals, Astronomy, Earth Science...
Government Elections, Military, Law, Taxes...	Social Science Languages, Archaeology, Psychology...
Health Disease, Drugs, Fitness, Nutrition...	Society & Culture Sexuality, Religion, Food & Drink...
New Additions 4/13, 4/12, 4/11, 4/10, 4/9...	Subscribe via RSS Arts, Music, Sports, TV, more...

April 13, 2013

Historical Trends: Search on Taxonomies

Search Terms

Search for

Terms

Field

in Subject

Gale OneFile (access via your library)

Subject

☐ ECONOMIC GROWTH 7.5K

☐ INCOME 6.6K

☐ PRIVATE SECTOR 5.1K

☐ ECONOMIC DEVELOPMENT 4.6K

☐ TRANSPARENCY 4.6K

[Show more](#)

Search Subject

Subject

openknowledge.worldbank.org

Scenarios for Information Seeking

When do people tend to search vs. browse vs. query?

- Search



- For finding something

- Finding content items (documents, pages, images) on a specific topic. Traditional topic-based search.
 - Finding *something* specific: a document, an image, a video, a sound recording, a product, a service

- Browse



- For exploring or seeking to discover something

An image or digital asset *suitable* for a use, a gift for someone, media or entertainment experience (with more control than recommendations)

- Query



- For getting information

An answer to a question (simple or complex), inferencing from data, or doing research

Search on Taxonomy Terms

Type ahead or auto-complete

Terms

marketing

Marketing

Marketing industry

Marketing executives

Marketing consulting services

Marketing management

Marketing communications

Marketing consultants

Marketing management software

Marketing personnel

Marketing cooperatives

Marketing models

Auto-suggest

Terms

taxonomy

Taxonomy (Biology)

Taxonomy of Educational Objectives (N

Taxonomy (Web browser)

Taxonomy

Animal taxonomy

Plant taxonomy

Taxonomía (Biología)

Taxonomía animal

Search on Taxonomy Terms

Search Types

- **Exact** - exact match
- **Contains** - exact match phrase with additional words before or after
- **Begins** - alphabetical from start, but allows end truncation (type-ahead)
- **Smart** - words within the term in any order and internal word stemming

Offering users such options is more common in library systems.

Find items that have:

All these words:

This exact phrase:

in the title	▼	contains	▼	(Don
in subject	▼	is (exact)	▼	Amu

Browsing: ERIC -- Thesaurus

☒ Term Begins With ☐ Term Contains ☐ Relevancy Ranked

Search on Synonyms (Alternative Labels) of Concepts

Searches may match on alternative labels of concepts, too.

Demo example from [PoolParty Help](#)

Search matches to dropdown terms:

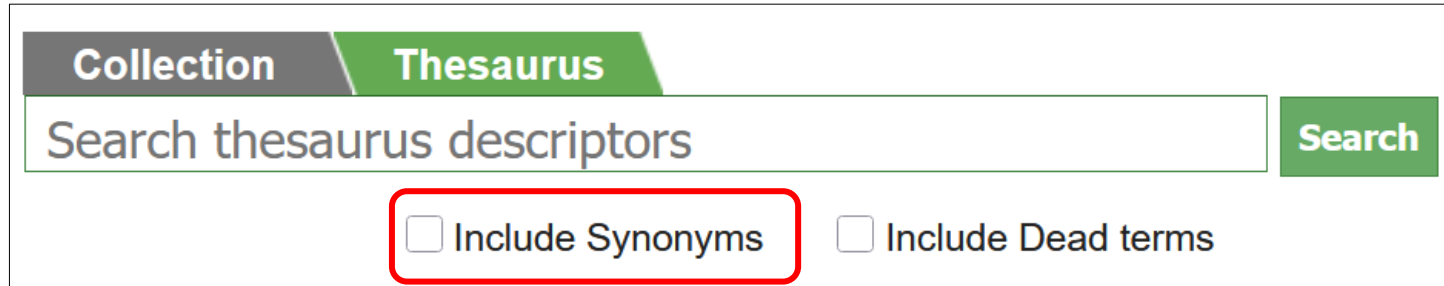
CSV import column format

Excel import format

Spreadsheet import format

Are all alternative labels for the concept: *Spreadsheet import format*

[ERIC Institute of Education Sciences thesaurus](#)



Collection Thesaurus

Search thesaurus descriptors Search

☒ Include Synonyms ☐ Include Dead terms

[Gov.uk demo link](#)

Search on Taxonomy Terms in Context

Terms in the context of concept scheme, facet, or metadata property

search|

GraphSearch

Context: PoolParty Product

Recommendation search engine

Context: Technologies and Applications

Elasticsearch

Context: Technologies and Applications

Search Space

Context: PoolParty Product

help.poolparty.biz

openknowledge.worldbank.org

Search the OKR ...

sustainable|

COLLECTION

Inspection Panel Emerging Lessons Series

Environment and Sustainable Development

Sustainable Infrastructure

AUTHOR

Global Platform for Sustainable Cities10

Sustainable Banking Network1

Sustainable Mobility for All1

SUBJECT

SUSTAINABLE DEVELOPMENT1904

SUSTAINABLE GROWTH605

SUSTAINABLE ENERGY338

SUSTAINABLE USE267

SUSTAINABLE MANAGEMENT258

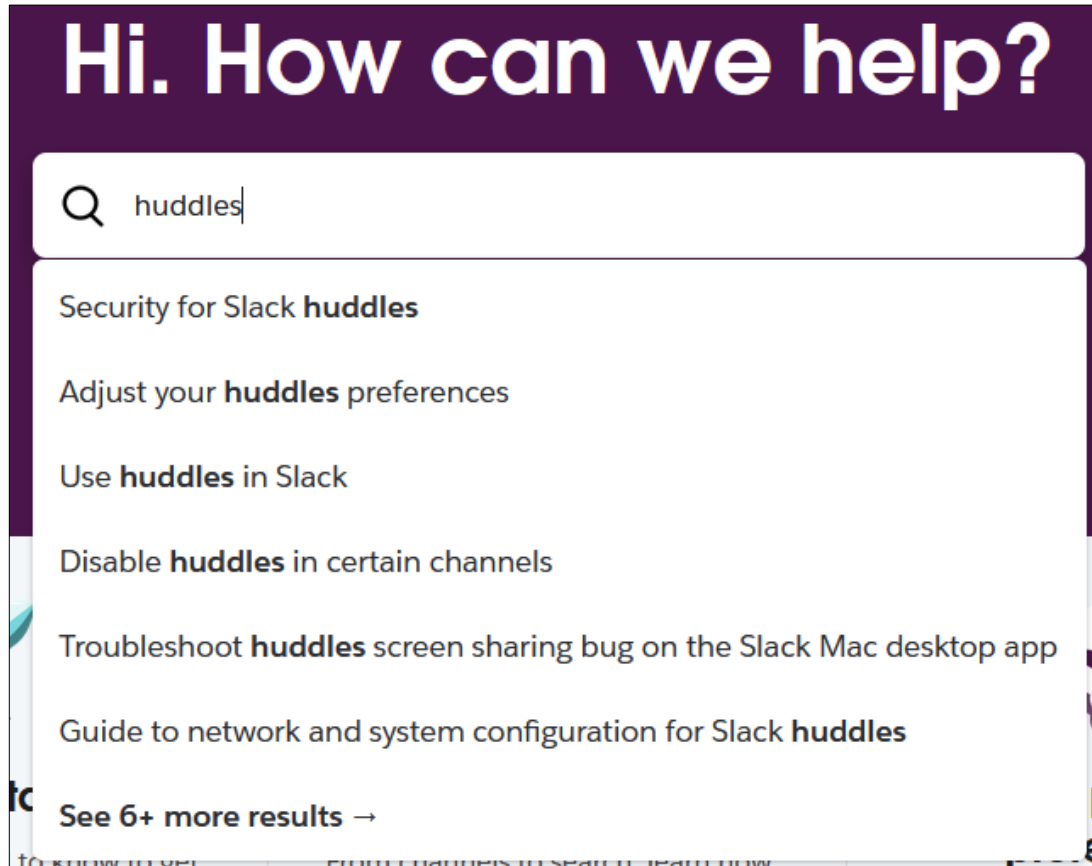
SUSTAINABLE FOREST MANAGEMENT149

JOURNAL

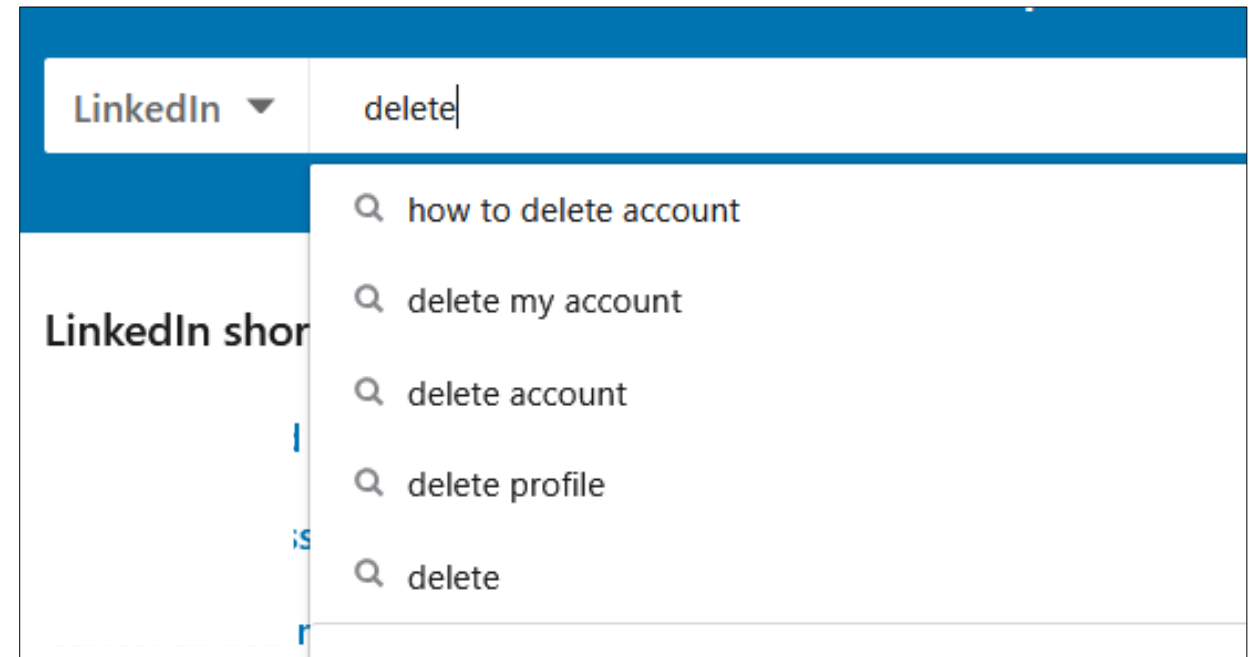
World Bank Economic Review

Search *Not* on Taxonomy Terms

Not all implementations of search suggest or auto-suggest involve taxonomies.



Search terms match to Help article *titles*.



Search terms match past searches.

Faceted Taxonomies as Search Refinements



“Facets” mean “face”: side, dimension, aspect

- Different aspects to be used in combination
- Facets could be for question word aspects: What, Where, When, Who, Why, How
- Each facet is managed in the SKOS data model as a Concept Scheme
- Each facet tends to be a metadata property and can serve as search filter.

Examples of taxonomy facets:

- *for market research:* Topic, Industry, Product Type, Brand Name, Company
- *for an Intranet:* Department, Region, Topic, Activity, Person Name, Event
- *for training content:* Skill, Role, Level, Subject, Delivery Method
- *for technical documentation:* User Type, Content Type, Product, Feature, Topic

“Facets” tend to refer to the subset of filters that are managed as controlled vocabularies

- Other metadata are not “taxonomies”: date, creator, price, availability/status, etc.

Update Search

< Refine Search

Resources (6,757,017)

Learning Lab Collections (12,588)

Found 12,588 Learning Lab Collections

Faceted taxonomies can be implemented for:

- Faceted browse: for initial for browsing
- Faceted search: for refining/limiting/filtering search results

Subject

- ☐ Social Studies (4353)
 - ☐ US History (2156)
 - ☐ Cultures (1347)
 - ☐ World History (787)
 - ☐ Civics (692)
 - ☐ Geography (478)
 - ☐ (262)
- ☐ Arts (2852)
 - ☐ Visual Arts (1374)
 - ☐ (501)
 - ☐ Music (234)
 - ☐ Performing Arts (196)
- ☐ Language Arts And English (2088)
 - ☐ Speaking and Listening (397)
 - ☐ (391)
 - ☐ Writing (373)

Age Range

Educational Features

Collection Type



The Birth

#ArtsIntegr

Created by:
Arts

The Gre

#ArtsIntegration

By Jennifer Mellizo
Arts Integration

Masks, Movement and Music: Journey through Nō (能) Theater

#ArtsIntegration

By Destini Collins
Arts Integration

Mythos, Music, and Movement in Ancient Ch

#ArtsIntegration

20

56

Echoes of the Drum:
Discovering Taiko
through Art and
Sound

learninglab.si.edu/search/

Attributes as Search Refinements

Attributes are refinement filters on search results

- Based on additional (data type) properties metadata for taxonomy concepts or categories.
- Often used in ecommerce, hospitality offerings, media offerings, etc.
- May vary depending on the taxonomy category
- In addition to controlled vocabularies may be numbers, dates, free text, etc.
- May be managed as ontology (RDFS) data type properties, rather than as SKOS concepts in concept schemes.

Examples of attribute metadata:

- *for people:* birth date, gender, occupation, nationality, phone number
- *for products:* brand, price, color, size, SKU number
- *for documents:* title, author, publication date, language, word count, file type

Brand

- ☐ Lenovo
- ☐ HP
- ☐ Dell
- ☐ ASUS
- ☐ Microsoft
- ☐ Samsung
- ☐ LG
- ☐ Acer

[Show all \(16\)](#)

Screen Size

- ☐ 10" - 11.9"
- ☐ 12" - 13.9"
- ☐ 14" - 15.9"
- ☐ 16" or More

RAM

- ☐ 128 gigabytes
- ☐ 96 gigabytes
- ☐ 64 gigabytes
- ☐ 32.0 gigabytes
- ☐ 32 gigabytes
- ☐ 24 gigabytes
- ☐ 16.0 gigabytes
- ☐ 16 gigabytes

[Show all \(11\)](#)

Processor Model

Post-Search Refinement

- Search results may be refined with taxonomy terms, without it being a faceted taxonomy.
- Taxonomy terms, along with other kinds of filters, are used to limit results.
- Implemented especially when a taxonomy or thesaurus is large, with too many concepts to browse through within a facet.

GALE

Academic OneFile

"Internet of things"

Basic Search

Advanced

Bookmark

More

Browse by Discipline

Magazines (3,776)

Search Terms: Basic Search ("Internet of things")

Sort by: Relevance

Citation: ARTICLE
NO BATTERIES INCLUDED: New forms of energy harvesting and low-powered sensors promise to eliminate batteries and enable the Internet of Things
Amit Katwala.
Professional Engineering Magazine.
30.9 (Oct. 2017) p24+. Word Count: 1990. Reading Level (Lexile): 1360.
Link to CDL Materials Catalog ILL My Library OpenURL Link
Save

Citation: BRIEF ARTICLE
INTERNET OF THINGS IS 'COMING OF AGE'
Amit Katwala.
Professional Engineering Magazine.
30.9 (Oct. 2017) p6. Word Count: 217. Reading Level (Lexile): 1110.
Link to CDL Materials Catalog ILL My Library OpenURL Link
Save

Full-text: ARTICLE
The internet of things that can be hacked
Modern Healthcare.
48.4 (Jan. 22, 2018) p0022. Word Count: 1493. Reading Level (Lexile): 1290.
Byline: Rachel Z. Arndt When doctors replaced then-Vice President Dick Cheney's pacemaker in 2007, they asked the

Content Types

Academic Journals (6,397)

Magazines (3,776)

Books (24)

News (38,942)

Images (2)

Videos (4)

Search within results

Limit Search by:

☐ Full Text

☐ Contains Images

Publication Dates:

Past Week

Past Month

Past Year

Custom Date Range

Subjects

Internet of things (724)

Software industry (197)

Semiconductor indust... (120)

Creating Synonyms for Concepts to Support Search

Alternative labels / variants / synonyms help match user search strings to strings of text in the content via controlled vocabulary (taxonomy) concepts.

- Synonyms to terms in the texts support more comprehensive tagging.
- Synonyms to user-entered search strings support more comprehensive retrieval.

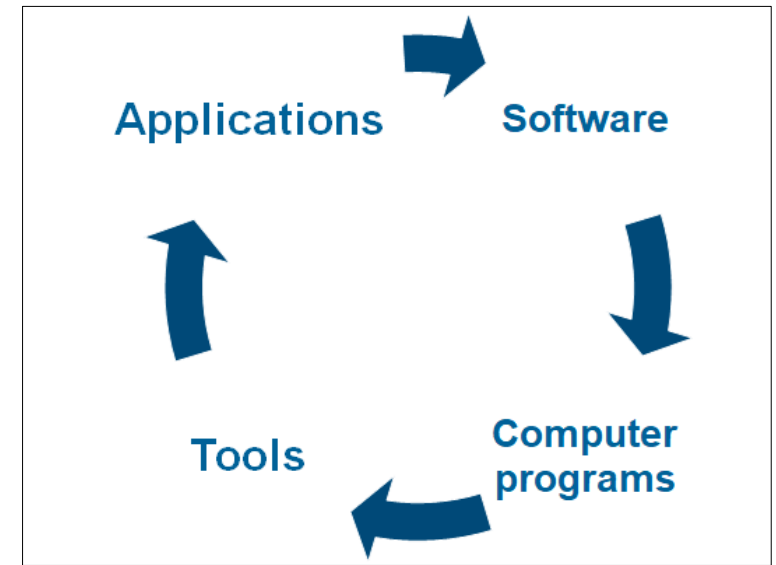
Example:

Car industry can be an alternative label for Motor vehicle industry in tagging.

- Content mentioning the car industry is about the Motor vehicle industry.
- Car industry cannot be an alternative label to Motor vehicle industry in search/retrieval.
 - A search on car industry may retrieve content about the truck industry, which was correctly tagged with Motor vehicle industry.
- Do *not* designate narrower terms for concepts as synonyms for search, although you can for tagging.
- Instead, have a separate “search taxonomy,” a synonym ring/search thesaurus, or designate different kinds of alternative labels for a concept (such as using SKOS Hidden Label or SKOS-XL)

Search Thesauri (Synonym Rings)

- A simple controlled vocabularies with two or more synonyms for a concept, but no preferred displayed label.
- Instead of an entire separate search “taxonomy,” contains just the list of terms for search
- Specifically, for matching end-user search strings to text strings in the content being searched, for commonly searched concepts by various names.
- Since synonyms are not displayed to the users, common misspellings, slang, deprecated, and potentially offensive terms may be included.
- May be the only controlled vocabulary implemented, or implemented in addition to a taxonomy use for tagging and browsing.
 - In which some of the synonyms may need to be repeated in the other taxonomy for tagging.



Search Thesauri (Synonym Rings)

Types of “synonyms” for search thesauri

- Synonyms: Cars / Automobiles
- Quasi-synonyms in common phrases: learning resources / training resource
- Variant English spellings, especially between US vs. UK spellings: center, centre
- Former and names: Bangalore / Bengaluru
- Acronyms and spelled out form: SEC / Securities and Exchange Commission
- Abbreviations and spelled out form: pharma / pharmaceutical
- Nonstandard pluralization that don’t get resolved by typical stemming: mouse / mice
- Variations of using a space: data mart / datamart
- Use of special characters or not: 401(k), 401k

1	Synonym 1	Synonym 2	Synonym 3
819	geese	goose	
820	GenAI	Gen AI	Generative AI
821	genera	genus	
822	generalisation	generalization	

Tagging Synonyms and Rules for Auto-tagging

Synonyms/alternative labels for auto-tagging need to:

- Anticipate a greater number of variations that may appear in texts than needed for manual tagging or for end-user browsing/drop-down
- Disambiguate different meanings of the same word or name.

Including Boolean logic-based rules is a method of auto-tagging.

- Can substitute for numerous alternative labels with proximity operators
- Can incorporate disambiguation with NOT operator

Rule to find in text: **transportation near:4 (regulation or regulations or law or laws or policy or legislation)**

To tag with the concept: **Transportation regulation**

Rule to find in text: **Acerta near:15 (automotive or (manufacturing near:3 (platform or software)))**

To tag with the concept: **Acerta Analytics**

Rule to find in text: **BrightPath not (solar or Solar)**

To tag with the concept: **BrightPath Early Learning**

Taxonomies in Support of Querying

Standard Chatbots and Question Answering Systems

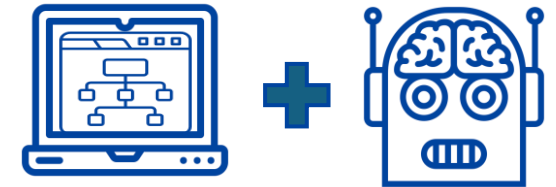


- Users ask question in natural language to get answers.
- Natural language processing (NLP) parses questions, which are matched to taxonomy terms that have been tagged to answers.

LLM-based Generative AI systems

Taxonomies can provide context for language models and thus more accurate results, especially for LLMs used within an enterprise.

- Context of the subject of the content being queried
- Context of the subject of the content used for training the model



RAG (Retrieval Augmented Generation), supported by taxonomies by:

- structuring the knowledge that RAG retrieves.
- providing RAG systems with domain-specific terminology
- ensuring that the retrieved content is both topically related but semantically correct

Scenarios for Information Seeking with Taxonomies

Search types vary:

(for enterprise, intranet, website CMS, not the Web)

- **For a certain document:** Search on title may be good, but better with facets on source, date, etc.
- **For information on a topic:** A topical taxonomy is best.
- **For complex queries:** A taxonomy + ontology with semantic relations
- **For data:** A taxonomy + ontology in a knowledge graph (graph data)
- **How to do something:** A taxonomy for a category, and good text analytics/NLP based search
- **Other questions:** A taxonomy combined with a QA system or GenAI + LLMs if the repository is trained, enhanced with RAG

Further Reading on Taxonomies and Search

The Accidental Taxonomist Blog posts:

[Topical Taxonomies for Filtering Searches](#), September 2024

[Synonym Rings \(or Search Thesaurus\)](#), April 2024

[Faceted Search vs. Faceted Browse](#), April 2012

Presentation slides:

[Indexes, Search, and Taxonomies: Paths to Findability](#), January 2025

[Semantic Search: Implementations for Your Enterprise](#), November 2022

[Taxonomies in Support of Search](#), October 2018

General Taxonomy Resources

- ANSI/NISO Z39.19-2005 (2010) Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies
www.niso.org/publications/ansiniso-z3919-2005-r2010
- The Accidental Taxonomist Blog
<http://accidental-taxonomist.blogspot.com>
- Accidental Taxonomist book websites
www.hedden-information.com/accidental-taxonomist/websites
- Hedden Information Management past presentations
www.hedden-information.com/presentations
- Hedden Information Management taxonomy training
www.hedden-information.com/courses-workshops
- Taxonomy Talk, taxonomists community on Discord
<https://discord.com/invite/3qyMVYCAsw>

Questions/Contact

Heather Hedden

Taxonomy Consultant

Hedden Information Management

Carlisle, MA

heather@hedden.net

www.hedden-information.com

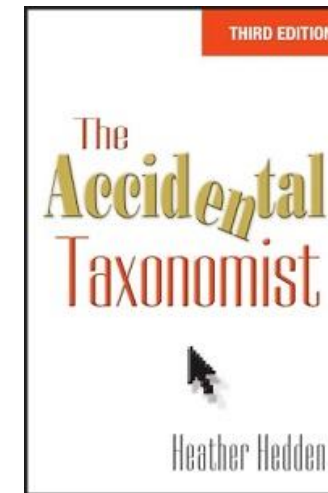
accidental-taxonomist.blogspot.com

www.linkedin.com/in/hedden



**Hedden Information
Management**

Making information findable



For sale at the Information Today Inc. book publisher booth.
Book signings today 5:00 – 6:00 pm and Tuesday 5:00- 5:30 pm